Guangxu Jin, Kemi Cui, Xiaobo Zhou,
and Stephen T.C. Wong

Unraveling the Signal-Transduction Networks in Cancer Metastasis

Metastasis is a major cause of morbidity and mortality in cancer patients. For example, in the United States, the five-year relative survival in local prostate cancer patients is nearly 100% while it is only 32% in patients with distant metastases (http://www.cancer.org). Similar to many types of malignant tumors, prostate cancer develops from the mutations or disregulation of various genes involved in regulation signal transduction and the cell cycle [1]. The discovery of molecular pathways that significantly contribute to tumor metastasis would provide specific targets and optimal strategies in prevention and treatment of such malignant diseases.

We describe a new approach for revealing the signal-transduction networks based on the clustering of network topologies, multiple genomic expression profiles, and cell signaling pathways and illustrate its utility by applying it to the analysis of prostate cancer metastasis. The resulting signal-transduction networks composed of those protein paths not only achieve high differential expression signatures in multiple microarray data sets but also function as the signal-transduction bridges linking the prostate cancer signaling pathway and other cell signaling pathways. The identified networks assemble both network topology and expression signatures for prostate metastasis into a meaningful biological context through signaling pathways and have the potential to be the key signal transductions for prostate cancer metastasis. By validating the identified networks using information obtained from molecular function (GO), tissue expressions, patients classification, Drugbank, and published results, we found that they not only shared the same molecular functions, are highly expressed in prostate tissue, and achieve high-accuracies in stratifying prostate cancer patients but also are targets for the drugs used in treatments of prostate cancer metastasis and are important ingredients of prostate cancer cell migration and metastasis.

Digital Object Identifier 10.1109/MSP.2009.933384

THE SIGNAL-TRANSDUCTION NETWORK DISCOVERY

Our discovery of signal-transduction networks for prostate cancer metastasis consists of the following three steps: the protein-paths identification, protein-paths selection, and validation. Initially, the protein paths were identified from both clustering network topologies between signaling pathways and multiple genomic expression profiles. The available signaling pathways are indispensable in the protein-paths identification. They ensemble the network topology of protein network into biological context so that the selected protein-paths from the clustering network topologies can be identified as the key signal transductions connecting different signaling pathways. Moreover, after mapping between proteins and genes, multiple genomic expression profiles were used to demonstrate differential expression levels of genes in the selected protein-paths. To select the high-confident paths from those identified protein-paths, we then proposed a certain differential score for the selected paths, i.e., differential expression score (DES) of paths (DESP). DESP was computed from the DESs of the composite proteins in the particular path. The paths with highest DESPs were identified and defined to be related to prostate cancer metastasis. Last, the identified paths were validated with respect to known molecular functions, tissue expression, patient classification, drug targets, and published results.

THE CLUSTERING NETWORK TOPOLOGIES LINKING THE SIGNALING PATHWAY OF PROSTATE CANCER

The approach to determine the signal transductions specific to cancer metastases is to build up the signal-transduction bridges between the prostate cancer signaling pathway and other cell signaling pathways. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database for general signaling pathways, whereas KEGG DISEASE is the database for the signaling pathways specific to human diseases [2]. There are two challenges for using the available signaling pathways to explain the tumor metastases. One is that the publicly available cancer signaling pathways can only provide general information of signal transductions for cancers but are not sufficient to interpret special cases of cancers, such as cancer metastases. The other is that cancer metastases is a complex biological process of heterogeneous cells and involves multiple organs, which suggests that multiple cell signaling pathways might be involved. Thus, it is necessary to identify more specific cancer-metastasis-related genes to build up the bridges between signaling pathways so that complete signaling mechanisms for special cases of cancers can be revealed.

The identification of cancer-metastasis genes is performed by both clustering network topology in protein network and the DES defined by multiple genomic expression profiles. The clustering network topology in protein networks will be first presented before the introduction of DES.

Since the conserved patterns, i.e., network motifs [3], [4], play important roles in signal transductions, we evaluated their topological properties with respect to protein network and signaling pathways. The network motifs display their clustering property in the protein network. To indicate the contribution of the clustering network topology to signal transductions, we proposed a novel quantitative criterion, i.e., clustering *P*-value. The clustering *P*-value for signaling pathways was computed from the complementary cumulative distribution function (ccdf) derived from a random process. As illustrated in Figure 1, certain clustering network motifs are specifically clustered between the prostate cancer and cell cycle pathways evidenced by the clustering *P*-value for prostate cancer and cell cycle signaling pathways being significantly low ($P < 0.01$).

THE DES DEFINED FROM MULTIPLE MRNA EXPRESSION PROFILES

To identify the cancer-metastasis-related genes, the DESs for the genes were established after the identification of their mapping proteins acting as the connection bridges between signaling pathways. To avoid the negative effects of data noises in microarray data sets on the identification of the cancer-metastasis-related genes, we used the available microarray data sets for prostate cancer metastasis (ten microarray data sets from Oncomine [5] and NCBI GEO [6] instead of single microarray to define the DESs).

For each gene j, we proposed an integrated *P*-value, i.e., DES, based on all of these ten microarray data sets

$$\mathrm{DES}_j = -\sum_{i=1}^{10} \log(p_{ij})$$

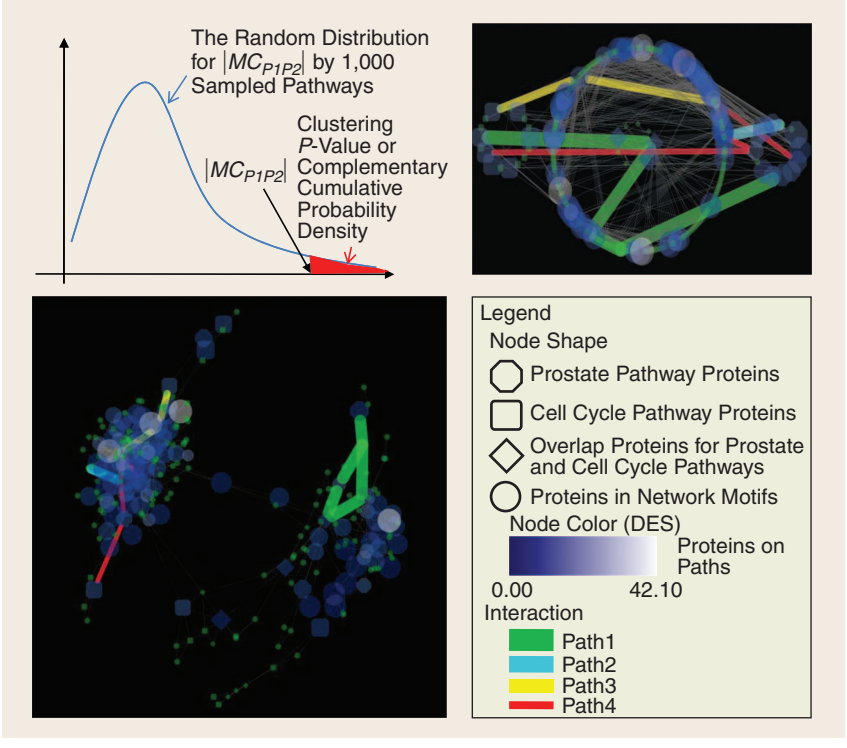
where P_{ij} is the P value to evaluate the differential expressions of the gene j in the two different conditions of the ith microarray data set (student t test).

For a path k chosen from the clustering network motifs, we derived a combined *P*-value, DESP, by combining the corresponding DESs of the composite proteins,

$$\mathrm{DESP}_K = \frac{1}{K}\sum_{j=1}^{K} \mathrm{DES}_{jk}$$

where K is the number of proteins in the path k.

By combining the differential *P*-values of a gene in multiple microarray data sets together, the negative effects stemmed from relatively small number of microarray data sets were reduced. The DES derived from multiple microarray data sets instead of single microarray can significantly reduce the error occurrence caused by data noises presented in microarray data sets.



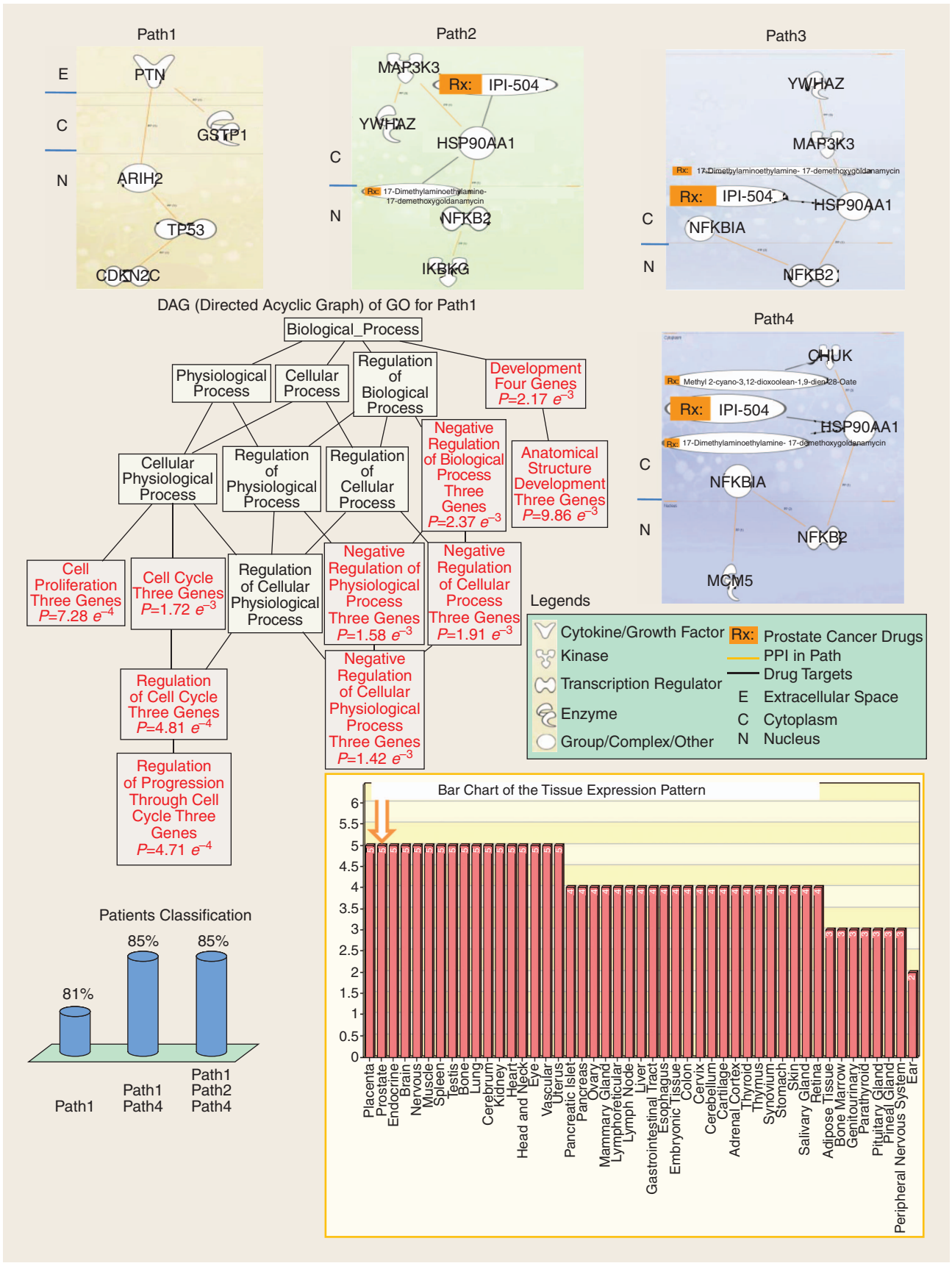
[FIG1] The clustering network motifs between prostate cancer and cell cycle.

THE HIGH DESP PATHS

The signal-transduction paths for prostate cancer metastasis were identified from the clustering network topology between the prostate cancer and cell cycle signaling pathways and the defined DESs from multiple genomic expression profiles. To indicate the signal transductions between the two signaling pathways, we chose the protein paths acting as the connection bridges between the two signaling pathways in the clustering network topology. The DESs of genes were then defined by combining multiple microarray data sets, which were used to choose those high-confident paths for prostate cancer metastasis. The DESP for a path was defined as the mean value of DESs in the path. It has the ability to indicate the expression difference of the paths between prostate carcinoma and prostate carcinoma metastasis patients. Finally, four high-confident paths with the highest DESPs were identified; the four paths were called path1, path2, path3, and path4.

THE BIOLOGICAL MEANINGS OF CHOSEN PATHS

After the identification of paths from clustering network topology and multiple microarray data sets, we examined their biological meanings in six facets, i.e., biomarkers and drugable targets, GO molecular function, tissue expression, cellular signal transduction, patients classification, and published results. First of all, a number of genes and proteins in the identified four paths are targeted by three selected drugs used in the treatments of prostate cancer whose targets are HSP90AA1 and CHUK (Figure 2). Some of genes in the found paths have been suggested as possible diagnostic and prognostic markers for metastatic prostate cancer, such as p53, MCM5 (minichromosome maintenance complex



[FIG2] The identified four high-confidence paths and their validations.

component 5), GSTP1, and NFκB. Potential targets for malignant prostate cancer, such as PTN (pleiotrophin), MAP3k3 (mitogen-activated protein kinase) NFκB, and CHUK (conserved helix-loop-helix ubiquitous kinase) are also included in the selected paths. More importantly, it has been reported that PTN in path1 is related to prostate cancer migration and NFkBIA in path3, and path4 is associated with prostate cancer metastasis in the publications. Such collected information indicates that the selected paths are valuable for biomarkers and targets discovery.

Second, by checking the shared GO terms of the proteins in the identified paths, we derived that most proteins in the paths share the same GO terms ($P < 10^{-2}$) [7], which suggests that the proteins on the selected paths may function similarly (Figure 2). The GO molecular function information indicates that most proteins in the identified paths are responsible for the similar tasks related to development of prostate cancer. Many proteins in the identified paths are highly expressed in the prostate tissue. Moreover, the information of cellular localization suggests that the proteins in the path are responsible for the signal transductions between cross-talking of cytoplasm and nucleus (Figure 2).

Last, the paths display high accuracies in patient classification (as high as 85%) as shown in Figure 2. The features that are input into a support vector machine (SVM) are the combined path expressions computed by the corresponding gene expressions of composite proteins in the path. After mapping the gene expressions from single proteins to paths, we identified the potential of the paths to serve as the biomarkers for classifying metastasis patients. The results of the patients classification indicate that the selected paths may predict the possibility of prostate cancer migrating to the other tissues. The validations from available biological information indicate that the four selected paths are of high confidence in the central roles of prostate metastasis, and our approach of such biological signal processing is promising to in discovering diagnosis and prognosis biomarkers as well as treatment targets.

ACKNOWLEDGMENTS

This research is funded by NIH R01LM08696, R01LM009161, and R01AG028928.

AUTHORS

Guangxu Jin (Gjin@tmhs.org) is a research fellow in the Center for Biotechnology and Informatics, The Methodist Hospital Research Institute.

Kemi Cui (Kcui@tmhs.org) is a research associate in the Medical Systems Biology Laboratory, the Center for Biotechnology and Informatics, The Methodist Hospital Research Institute.

Xiaobo Zhou (Xzhou@tmhs.org) is an associate professor of radiology, Weill Cornell Medical College, and chief of the Bioinformatics Laboratory, the Center for Biotechnology and Informatics, The Methodist Hospital Research Institute.

Stephen T.C. Wong (STWong@tmhs.org) is the John S. Dunn Distinguished Endowed Chair of Biomedical Engineering, professor of radiology at Cornell University, and director of the Center for Biotechnology and Informatics, The Methodist Hospital Research Institute.

REFERENCES

[1] J. T. Lee, B. D. Lehmann, D. M. Terrian, W. H. Chappell, F. Stivala, M. Libra, A. M. Martelli, L. S. Steelman, and J. A. McCubrey, "Targeting prostate cancer based on signal transduction and cell cycle pathways," *Cell Cycle,* vol. 7, no. 12, pp. 1745–1762, June 2008.

[2] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi, "KEGG for linking genomes to life and the environment," *Nucleic Acids Res.,* vol. 36, pp. D480–D484, Jan. 2008.

[3] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science,* vol. 298, no. 5594, pp. 824–827, Oct. 2002.

[4] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nat. Genet.,* vol. 31, no. 1, pp. 64–68, May 2002.

[5] D. R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, and A. M. Chinnaiyan, "ONCOMINE: A cancer microarray database and integrated data-mining platform," *Neoplasia,* vol. 6, no. 1, pp. 1–6, Jan./Feb. 2004.

[6] T. Barrett, T. O. Suzek, D. B. Troup, S. E. Wilhite, W. C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi, and R. Edgar, "NCBI GEO: Mining millions of expression profiles—Database and tools," *Nucleic Acids Res.,* vol. 33, pp. D562–D566, Jan. 2005.

[7] B. Zhang, S. Kirov, and J. Snoddy, "WebGestalt: An integrated system for exploring gene sets in various biological contexts," *Nucleic Acids Res.,* vol. 33, pp. W741–W748, July 2005. [SP]

from the GUEST EDITORS continued from page 17

This tutorial provides an overview of coalitional game theory concepts and their applications in communications and wireless networks. The mathematical tools and techniques needed to study coalitional games are presented for three classes of games: canonical coalitional, coalition formation, and coalitional graph games.

The seventh article, "Natural Cooperation in Wireless Networks," by Yang, Klein, and Brown, discusses selfish behavior in networks on different layers using tools from noncooperative game theory. Natural cooperation without extrinsic incentive mechanisms is achieved in a repeated game framework, with credible punishments for defection and under long-term payoffs. The efficiency of this framework is illustrated by several relevant communication scenarios.

Finally, Scutari, Palomar, Pang, and Facchinei introduce a variation inequality (VI) framework for modeling and solving the interaction problems of rational entities in their article "Flexible Design of Cognitive Radio Wireless Systems." This framework integrates and supplements classical game theory and the article reflects the frontier of research in this area. The authors focus in particular on VI techniques for solving problems in the field of resource allocation (power, bit rate) in cognitive radio networks with variable interference constraints.

A NOTE OF THANKS

We received an overwhelming response to the call for this issue. We would like to thank all of the authors for their contributions. We also thank the reviewers for their help in selecting articles, and finally, many thanks go to Associate Editors Doug Williams and Dan Schonfeld for their support. [SP]